

## ONLINE STRATEGY INSTRUCTION FOR INTEGRATING DICTIONARY SKILLS AND LANGUAGE AWARENESS

**Jim Ranalli, Iowa State University**

This paper explores the feasibility of an automated, online form of L2 strategy instruction (SI) as an alternative to conventional, classroom-based forms that rely primarily on teachers. Feasibility was evaluated by studying the effectiveness, both actual and perceived, of a five-week, online SI course designed to teach web-based dictionary skills integrated with knowledge of lexical patterns (involving grammatical collocation, complementation, and transitivity). Sixty-four learners in a US university ESL composition course were matched for vocabulary size and then randomly assigned to treatment and comparison groups in a learning management system. The SI treatment comprised multimedia presentations and a variety of practice activities with immediate, specific feedback, while the comparison condition involved recurrent use of dictionaries for referencing vocabulary usage features but no instruction. Performance data showed significant gains among the SI group in contrast to the comparison group, as well as a clear need for such training, while perception data indicated that participants felt the instruction was effective and relevant. The findings point to the as yet untapped potential of this form of instruction to address long-standing cost-benefit concerns about SI, as well as to help learners make more strategic use of CALL resources.

**Keywords:** Language Learning Strategies, Vocabulary, Online Teaching and Learning, Multimedia

**APA Citation:** Ranalli, J. (2013). Online strategy instruction for integrating dictionary skills and language awareness. *Language Learning & Technology*, 17(2), 75–99. Retrieved from <http://llt.msu.edu/issues/june2013/ranalli.pdf>

**Received:** January 28, 2012; **Accepted:** October 26, 2012; **Published:** June 1, 2013

**Copyright:** © Jim Ranalli

### INTRODUCTION

Despite its decades-long history as a focus of scholarly interest, L2 strategy instruction (SI) has yet to fulfill its promise. While it is broadly accepted that achievement in L2 learning depends in large part on learners' own self-initiated and self-directed efforts, projects designed to teach helpful learning behaviors to those learners most in need of them have faced a number of challenges, not least of which are lingering questions about whether classroom-based SI is worth the effort and time. A recent meta-analysis by Plonsky (2011) found an overall small to medium effect for SI studies, with effectiveness moderated by several contextual, treatment, and outcome variables. Studies addressing certain strategy domains (e.g., reading, vocabulary) had larger effect sizes than others (e.g., grammar, listening), as did studies focusing on fewer rather than more strategies and those involving longer versus shorter treatments, which could "prompt some to question whether the benefits of SI outweigh the costs" (Plonsky, 2011, p. 23).

Teachers, to whom the burden of SI usually falls, may find that SI conflicts with curricular goals, their teaching beliefs or teaching styles, or the expectations of students and parents (Rubin, Chamot, Harris, & Anderson, 2007). Even when teachers are willing to undertake SI, they may be ill-equipped in terms of skills and knowledge for promoting strategies, instructional materials, and institutional support. In their seminal book on L2 strategies, O'Malley and Chamot (1990) identified teacher preparation for SI as the greatest obstacle facing the field. More recently, Chamot (2005) has discussed the need to investigate the degree to which effective SI is tied to the characteristics and experiences of individual teachers.

Another key factor in the cost-benefit calculus is the methodology used for SI. In their systematic review of 38 SI studies, Hassan et al. (2005) drew a contrast between interventions based on awareness-raising techniques versus those based on behavior-modeling approaches, and called for more research to determine the relationship between this variable and the durability of gains from SI. To the extent that any SI intervention involves skill acquisition and thus will require that learners be given sufficient practice opportunities (DeKeyser, 2007), ideally accompanied by immediate and specific feedback (Carlson, 2003), this constitutes yet another challenge for teacher-led forms of strategy instruction.

Finally, the difficulties of conducting SI research have been noted (e.g., Chamot, 2005; Hassan et al., 2005; Plonsky, 2011). Researchers must choose between laboratory settings, which allow for random assignment and control over instructional variables, or classrooms, which offer participants with real-world goals and motivations and thus greater ecological validity. The result, according to A. D. Cohen and Macaro (2007), is that “there is still much work to be done on strategy instruction in order to prove to learners, teachers, and the wider SLA research community that such an undertaking in the classroom is worthwhile” (p. 284).

But what if some elements of SI could be removed from the classroom and delivered through the computer, thereby lifting some of the burden from teachers and timetables? SI could thus emulate other areas of endeavor, such as corporate job skills, in which online instruction has been used cost-effectively to complement or even replace existing face-to-face training programs (Kim, Bonk, & Zeng, 2005). It could also allow SI to share in the many benefits of CALL already enjoyed by other areas of L2 instruction, including convenient access, easier distribution and recycling of materials, situated learning, the use of multimedia, data on learner performance and progress, immediate and user-specific feedback, and learner empowerment (Reinders & White, 2010). In addition, it could address recently discussed needs to help learners make more strategic and self-directed use of CALL resources (Hauck & Hampel, 2008; Hubbard, 2004; Winke & Goertler, 2008). Finally, through the capacities of learning management systems (LMSs), online SI might also facilitate more rigorous and ecologically valid research designs.

While a considerable amount of scholarship has investigated the use of L2 learning strategies in technology-mediated environments (e.g., Chapelle & Mizuno, 1989; Harris, 2003; Hauck & Hampel, 2008; Huang, Chern, & Lin, 2009; Jamieson & Chapelle, 1987; Li, 2009; Pujolá, 2002; Ulitsky, 2000), few studies to date have evaluated computers as a delivery mechanism for SI. Among those that have, A. D. Cohen and Ishihara (2005), O'Bryan and Hegelheimer (2007), and Sykes and A. D. Cohen (2008) were small-scale, preliminary studies without controls, while Kohler (2002) and Dreyer and Nel (2003) combined computer-based SI with face-to-face training. The present investigation was therefore designed to directly test the feasibility of an automated, online form of SI by evaluating the efficacy of a purpose-built, prototype course through a combination of experimental and perception data.

It is axiomatic that the strategies targeted by SI should be appropriate to the given context and intended beneficiaries. Furthermore, while some SI approaches address strategies in a general way, experts have called for greater focus on the strategies needed for specific tasks (A. D. Cohen, 1998, 2007; Oxford, Cho, Leung, & Kim, 2004; Wenden, 1995). The context of the present investigation was a tertiary-level ESL writing course for international students with a strong technology component; one class per week is held in a computer lab and most assignments are distributed and collected via an LMS. The students have been identified in a placement exam as needing additional instruction before they can register for the normally required composition courses, with the exam scoring guide noting that their writing exhibits, among other things, “word-choice errors that obscure meaning” and “serious and frequent errors in ... word form” (English Department, n.d.). The study thus sought to teach a specific form of strategic behavior involving online dictionaries employed in vocabulary usage tasks,<sup>1</sup> which will be described in the following sections.

## Dictionary Use among L2 Learners of English

Today's L2 learners of English are doubly lucky. First, they have at their disposal powerful lexical resources that have been specially written for them and informed by corpus research—advantages not shared by learners of many other L2s (Frankenberg-Garcia, 2011). In addition, most of the major publishers who produce these dictionaries have now made them freely available online where, for many purposes, they can be consulted more quickly and conveniently than the paper-based versions (Dziemianko, 2010).

However, lexicographical research shows that pedagogical dictionaries go widely under-exploited and are often misused. Even advanced learners tend to prefer bilingual to monolingual dictionaries and to use them mostly for referencing L1–L2 translation equivalents (Atkins & Varantola, 1998; Frankenberg-Garcia, 2005). Learners have trouble distinguishing correctly among senses of polysemous words (Chan, 2011; Nesi & Haill, 2002), and often consult only the initial or final entries (Nesi & Tan, 2011). They tend to ignore explicit usage information provided via symbols or codes, relying instead on dictionary examples for guidance (Bogaards & van der Kloot, 2002; Chan, 2012; Dziemianko, 2006), which can lead them astray, for example, when they use a word's semantic associations to deduce facts about its syntactic behavior.

Some consultation problems must be attributed to the way particular dictionaries organize and present information (Dziemianko, 2011; Tono, 2011), as well as limited attentional resources that learners may be unwilling to expend in the midst of other L2 tasks. Monolingual resources in particular impose additional processing burdens (Neubach & A. D. Cohen, 1988; Tono, 2001) that may discourage their use.

Research also shows that learners often lack a basic understanding of certain lexical features, thus undercutting the basis for dictionary consultation. Chan (2012) found that 31 advanced ESL learners at a Hong Kong university had trouble using dictionaries to identify and correct lexical errors involving complementation, transitivity, and prepositional collocation. Some participants indicated in think-alouds that they did not regard these issues as important for grammaticality, while others in a focus group acknowledged never reading usage information in dictionaries. Similarly, Frankenberg-Garcia (2011)—in a study involving 211 learners of English at a Portuguese university—elicited preferred reference types for a variety of vocabulary-use tasks. Most students selected a grammar book, as opposed to any type of dictionary, as the best reference for searching prepositional collocations, suggesting an inability to distinguish between these common lexical features and prepositions of time and place. In a study involving 40 adult learners in UK language schools, Nesi (2000) surmised that many participants were unable to exploit usage information in dictionary examples because “the concepts of transitivity and grammatical collocation were very poorly understood” (p. 115).

## Integrating Dictionary Skills with Knowledge of Lexical Patterns

Studies of pedagogical dictionary use often conclude with calls for learner training, but such training has tended to make particular dictionaries and their data categories the starting point, rather than authentic language-use problems of the type that should prompt consultation (Lew, 2011). Thus, more recent studies (Chan, 2011, 2012; Frankenberg-Garcia, 2011; Laufer, 2011) have proposed improving learners' language awareness as part of dictionary-skills training. This was the approach adopted in the present study, whose goal was to teach the strategic use of pedagogical dictionaries for addressing errors involving grammatical collocation, complementation, and transitivity, errors that research shows are common in ESL and EFL writing (Chan, 2010; Hemchua & Schmitt, 2006).

As a preliminary step, these linguistic features were subsumed into a single, pedagogically friendly category using the system of linguistic description known as *pattern grammar*, proposed by Hunston and colleagues (Hunston & Francis, 1998, 2000; Hunston, Francis, & Manning, 1997) as part of the COBUILD project. Pattern grammar unites grammatical collocation, complementation, and transitivity

under the umbrella term *pattern* and emphasizes the relationships between patterning and meaning. For example, when used with an *-ing* pattern, the verb *remember* can mean “to have a memory of something” (e.g., “I remember growing up in southern California”), but when used with a *to*-infinitive it can mean “to not forget to do something” (e.g., “Remember to lock the door when you leave”).

Specific patterns such as these may be familiar to many learners, but not as part of a wider linguistic category or concept, because they are typically taught in an ad hoc way, if at all (Hunston et al., 1997). And yet patterns can be seen to feature prominently in today’s pedagogical dictionaries. For example, the entry for the verb *warn* in the online version of the Longman Dictionary of Contemporary English



**Figure 1.** Entry for *warn*, with patterns in bold. Used with permission from the online version of the Longman Dictionary of Contemporary English ([www.ldoceonline.com](http://www.ldoceonline.com)).

(LDOCE) shown in Figure 1, highlights several patterns in bolded blue letters, organized according to the separate senses to which they correspond.

To make it easier to discuss the dictionary-skills and language-awareness aspects of the study, these were combined into a single construct called *strategic pattern knowledge*. This was operationalized as the ability to use a pedagogical dictionary to identify and correct errors involving lexical patterns. While imparting strategic pattern knowledge might not directly enhance the learners’ store of productive vocabulary, it could empower them to deal with pattern grammar more independently in their learning and use of English—for example, in writing and revising. It might also go some way toward addressing an insensitivity to collocational relationships that has been blamed for L2 learners’ poor knowledge of formulaic sequences (Granger, 1998; Laufer, 2011; Wray, 2002).

The present study thus took the form of an experiment contrasting an automated, online, SI condition with a comparison condition that involved practice with online dictionaries but no explicit instruction. In line with the task-based nature of the strategy instruction, the evaluation employed a task-based form of assessment administered as a pre-test and post-test. To further enhance validity, strategic pattern knowledge was assessed using both product and process data (Abraham & Vann, 1996), with the collection of the latter being facilitated by the computer-based nature of this inquiry. A matched randomized block design was used to control for vocabulary size because a moderate but significant correlation with treatment effects had been observed in a pilot study. Perception data were also elicited

because learner attitudes toward the usefulness and relevance of specific strategies can influence SI outcomes (Rees-Miller, 1993; Wenden, 1991).

The specific research questions this study asked were:

1. How is the development of strategic pattern knowledge affected by membership in an automated, online SI condition versus an online dictionary practice condition?
2. How do participants perceive the SI materials in terms of efficacy and instructional value?

## METHODS

### Overview

The study took place in Fall 2011, beginning in the second week of the 15-week semester, as illustrated in Figure 2. The researcher met with each of three different sections of the composition course on a regularly scheduled computer lab day to explain the purpose of the study and recruit participants. Students were told gift cards would be awarded to randomly selected participants at the end of the semester, but regardless of participation, all students were expected to complete the treatments as part of the course requirements. After being registered in the project LMS, the students were shown video instructions for the pre-test twice and then had 20 minutes to complete it (see [video link 1](#)). A brief, online biodata questionnaire was then assigned for homework. The following week, each section again met in the computer lab to complete an online version of the Vocabulary Levels Test (described below), whose results were used to determine matched pairs for random assignment to groups created in the LMS. One group was randomly selected to work through the SI materials during weeks 4–8 while the other was assigned the dictionary practice activity. In week 9, computer lab days were again used for the post-test (also preceded by a showing of the video instructions twice), which concluded the experiment phase of the study. The two groups then switched conditions for weeks 10–14 so that everyone could experience the SI materials. Perception data was gathered using short questionnaires administered during the SI treatment and a longer evaluation at the end of the semester (see descriptions below).

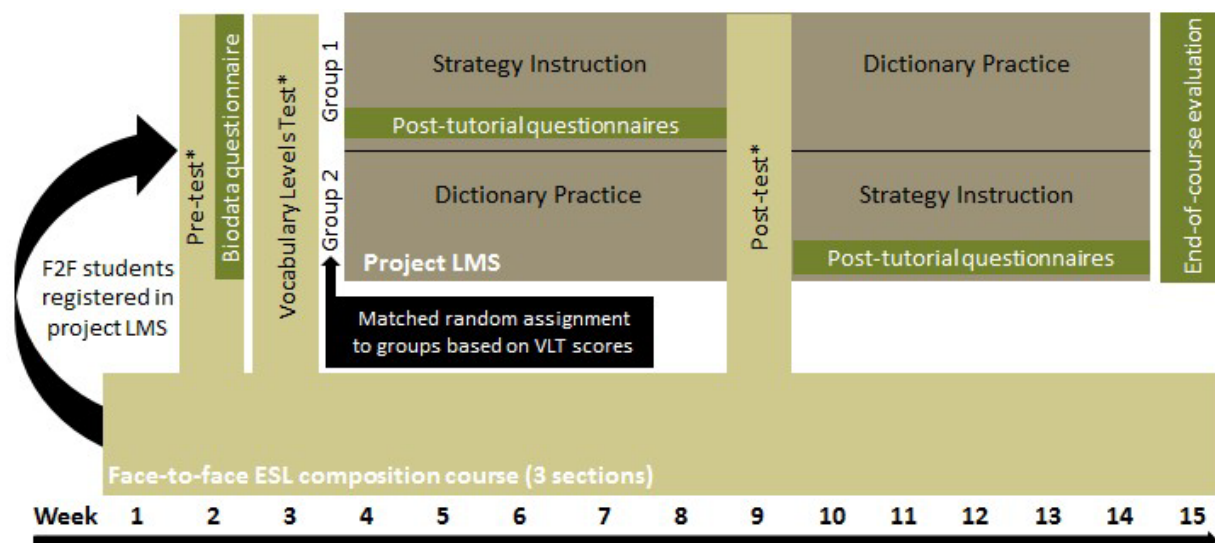


Figure 2. Timeline of procedures showing co-extension of LMS-based research activities with face-to-face composition course from which participants were recruited. Starred procedures (\*) were conducted in a computer lab.



## Participants

The sample ( $N = 64$ ) consisted mostly of students who spoke Mandarin Chinese as their first language (76.5%). While not very linguistically diverse, this sample was broadly representative of the current international student body at Iowa State University (according to 2011 figures) and other public universities across the United States, which is the population to which this study seeks to generalize. After matched random assignment to the strategy instruction (SI) group or dictionary practice (DP) group on the basis of vocabulary size, analysis showed the two groups did not differ significantly in this dimension.<sup>2</sup> As shown in Table 1, the resulting groups were also similar in terms of gender balance, variety of L1s, average time spent in the United States, age, and TOEFL score. Group size ( $n = 32$ ) was more than sufficient to meet the target ( $n = 6$ ) as determined by an a priori power analysis calculated in G\*Power (Faul, Erdfelder, Lang, & Buchner, 2007) using an effect size of 1.32 (Cohen's  $f$ ) obtained from pilot-study data, a desired power level of .95, and a significance level of .05.

Table 1. *Biodata for Treatment (SI) and Comparison (DP) Groups*

	SI group		DP group	
Sex	Female	14	Female	12
	Male	18	Male	20
L1	Chinese	26	Chinese	23
	Malay	2	Arabic	4
	[an African language]*	1	Korean	3
	Hindi	1	Malay	1
	Korean	1	Thai	1
	Spanish	1		
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Months in United States	7.5	14.0	5.7	8.4
Age	21.6	5.9	20.1	2.7
TOEFL score (iBT)	76.7	10.2	81.2	9.8

Note. SI = strategy instruction, DP = dictionary practice; \* Language not specified for reasons of confidentiality.

## Treatments

### *Strategy Instruction (SI)*

Participants in this condition were provided with strategy instruction in the use of learner dictionaries for identifying and correcting pattern-grammar errors. The instruction took shape in an automated, online course called VVT ("Virtual Vocabulary Trainer"), which lasted five weeks and consisted of eleven tutorials organized around three topics: vocabulary depth of knowledge (see discussion in Read, 2004), dictionaries, and pattern grammar. The first and third topics were given the more pedagogically friendly titles "Deep Vocabulary Knowledge" and "Word Patterns" respectively.

The course is based on instructional design principles that aim at (a) managing the demands made on learners' working memories to facilitate efficient learning, and (b) differentiating appropriately among instruction targeting domain-relevant declarative knowledge (e.g., conceptual understanding of pattern

grammar), procedural knowledge (e.g., syntactic processing of sentences to quickly identify patterns), or the successful integration and coordination of the two (see Ranalli, 2013, for a detailed description).

The tutorials consisted of video presentations and text-based practice activities designed to help learners:

- differentiate among dictionary types and choose the appropriate type for their purpose;
- use keyboard shortcuts and keyword search tools to speed up online dictionary searches, thus reducing demands on working memory;
- select the appropriate sense of a word from among those provided in dictionary entries;
- identify patterns in sentences and longer stretches of text;
- identify pattern information in dictionary entries;
- choose the appropriate pattern for a given context of use; and
- use pattern information in dictionaries to complete or correct sentences.

Additional activities were included to check learners' understanding of the material and review previously learned concepts and skills.

Figure 3 shows a screenshot of one of several video presentations aimed at elaborating learners' schema for pattern grammar by connecting examples they may already know, and then associating this provisional schema with the need for correctness in academic writing (see [video link 2](#)).

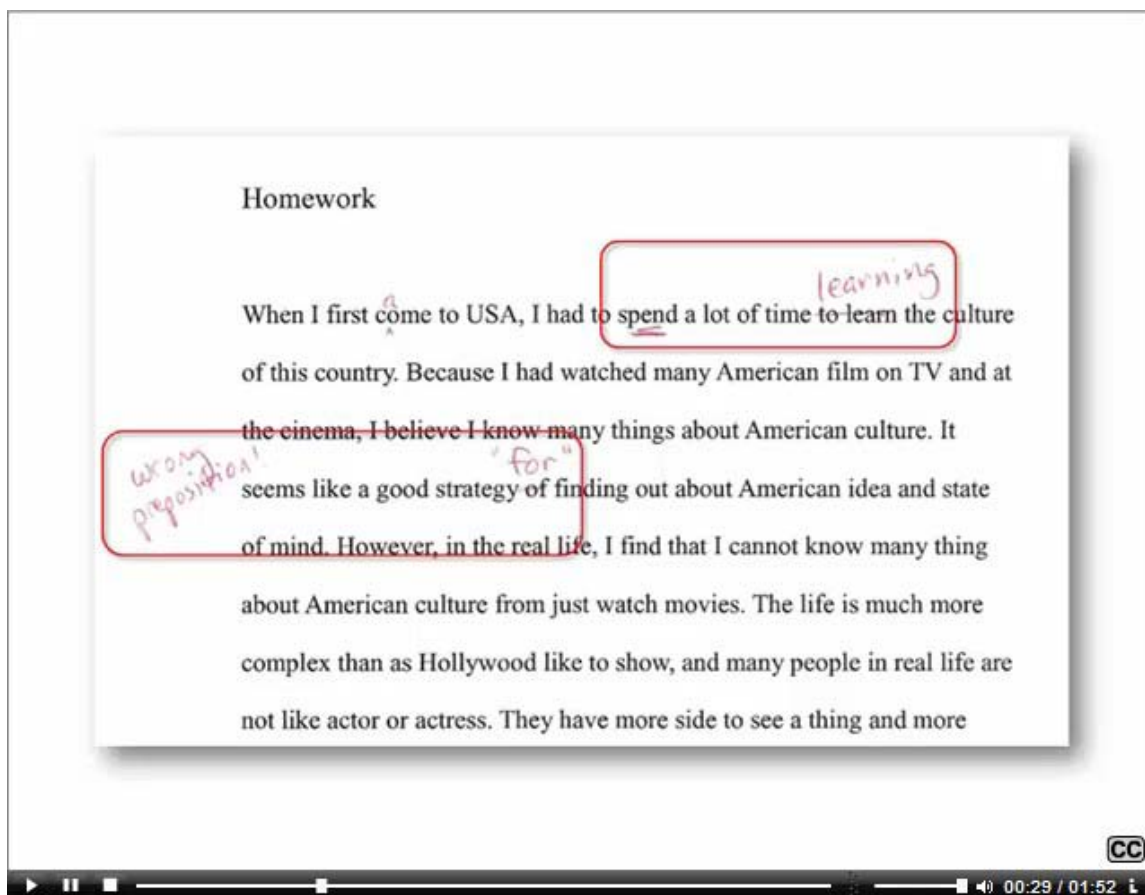


Figure 3. Screenshot of a video presentation from the SI treatment.

Figure 4 shows screenshots of a sequence of practice activities which further contribute to elaboration of a pattern-grammar schema. In the first exercise in Figure 4, learners study the abstracted patterns given at the top of the screen and then click on the instantiation of each pattern in the paragraph, which they locate with the help of their browser's "Find" tool. The latter follow-on activity features the same instantiated patterns now highlighted but missing components, which learners must then supply using dictionary data.

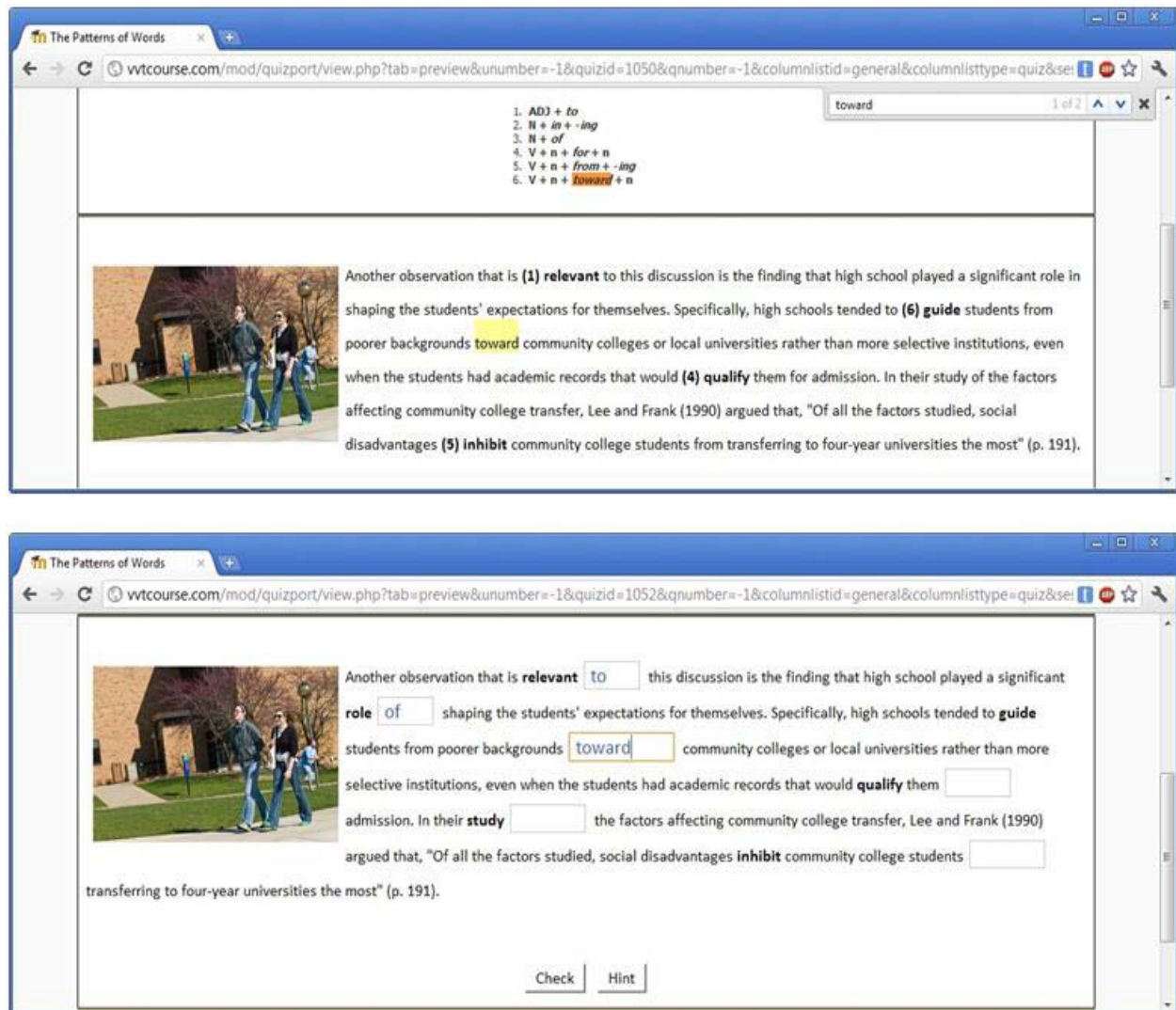


Figure 4. Part-task activities designed to elaborate learners' schema for pattern grammar.

### Dictionary Practice (DP)

Participants in this condition were assigned a recurring activity in the project LMS requiring them to add entries to a personal vocabulary database. The activity was an online version of the paper-based task that normally constitutes the vocabulary component of the composition course in question. Learners identify words in their course readings or other sources that they want to learn, then research and record lexical features and other information they think will help in using and remembering these words. The model provided to participants is shown in Figure 5. An instructional video accompanying each weekly assignment explained the purpose of the activity and demonstrated which dictionaries to use and how to create an entry (see video link 3). The database entries were graded each week, but the researcher gave no feedback regarding pattern grammar.



## Measures of Knowledge and Strategic Ability

### Vocabulary Levels Test (VLT)

Vocabulary size was measured using the Vocabulary Levels Test (Nation, 1990). A more recent version (Schmitt, Schmitt, & Clapham, 2001) was selected to reduce the chances that participants had previous


<b>WORD</b>	<b>innovation</b>	
<b>Source (book or website)</b>	<a href="http://edition.cnn.com/2009/BUSINESS/11/26/innovation.tips/index.html">http://edition.cnn.com/2009/BUSINESS/11/26/innovation.tips/index.html</a>	
<b>Original context (full sentence)</b>	<i>Innovation is not an inherent trait, it's a set of skills that anyone can learn.</i>	
<b>Part of Speech</b>	noun	
<b>No. of syllables - Stressed syllable</b>	4-3	
<b>Meaning 1</b>	a new idea, method, or invention	
<b>Meaning 2</b>	the introduction of new ideas or methods	
<b>Meaning 3</b>		
<b>Dictionary example sentence</b>	<i>Many people feel bewildered by the speed of technological innovation.</i>	
<b>Derivatives</b>	innovate (verb); innovative (adj)	
<b>Usage information</b>	innovation "in" something, e.g. "an innovation in web design"	
<b>Dictionary link</b>	<a href="http://www.ldoceonline.com/dictionary/innovation">http://www.ldoceonline.com/dictionary/innovation</a>	
<b>My own example sentence</b>	<i>In this new online course, we have introduced a number of exciting innovations that will make learning engaging and fun.</i>	
<b>Frequency list</b>	AWL	
<b>Submitted by</b>	Jim Ranalli	
<b>Week</b>	3	
<b>Date added</b>	Friday, 9 September 2011, 09:46 AM	

Image credit: Jose del Corral

Figure 5. Sample vocabulary database entry provided as a model to DP group participants.

experience of particular test items (see Appendix A). The test was adapted for online administration and included four frequency levels: the 2000- and 3000-word levels, which assess knowledge of high-frequency words; the 5000-word level, which represents a boundary between low- and high frequency words; and an academic word level, which tests knowledge of the Academic Word List (Coxhead, 1998). Each level contains a selection of 60 words, 30 of which are actually tested. The words are presented in six-item lists positioned next to three synonyms or glosses. Test-takers select a word from the list to best match each synonym/gloss by inputting the corresponding number into the appropriate gap (Appendix A).

One point is awarded for each correct response, and the scores are totaled for a maximum possible of 30 on each level and 120 on the test as a whole. Results of the reliability analysis are reported in Table 2.

Table 2. Reliability for the Online Version of the Vocabulary Levels Test

Level	Number of items	Cronbach's alpha
2000	30	.857
3000	30	.878
5000	30	.866
AWL	30	.923
Test as a whole	120	.964

Note.  $n = 64$

### Pattern Identification and Correction Task

Strategic pattern knowledge was assessed using an online task, which can best be described via example. Learners are presented with a web page in the LMS containing a sentence such as *In the 1950s, American scientists warned people for the harmful effects of television* (Figure 6). In both written and video instructions (see video link 1), they have been told the sentence contains a content word (i.e., verb, noun, adjective, or adverb) used incorrectly; specifically, there is a problem with the pattern of the content (or key) word that must be identified and corrected using an online dictionary; in this case, the verb *warn*. The learners are to copy and paste the sentence into an editing box, make an appropriate correction, and submit their answer. They are told not to change the key word in any way but to focus on the words around it, adding, deleting or changing them as needed but without altering the meaning of the sentence. A table of links to various dictionaries is provided.

In the sentence below, there is a problem with the **pattern** of one of the **content words**. (*Content words* include nouns, verbs, adjectives and adverbs.) You must identify which word is being used incorrectly, and then rewrite the sentence in the space provided to make it correct. You can COPY and PASTE the sentence into the editing box to save time. You should NOT change the content word itself in any way. Instead, focus on the words surrounding it. You may need to add, delete or change the surrounding words. Do NOT change the meaning of the sentence.

Use one of these dictionaries, or any other online dictionary, to help you. You may NOT use handheld electronic dictionaries/translators, paper dictionaries, mobile phones, or any resource besides online dictionaries.

In the 1950s, American scientists warned people for the harmful effects of television.

Answer:

Trebuchet

1 (8 pt)

Lang

**B** **I** **U** **S**

Path:

Figure 6. Item from pattern identification and correction task showing instructions, dictionary links, sentence to be corrected, and editing box.

While they could attempt a correction using existing knowledge, the sentences have been specially selected to contain lexical patterns that are likely to be unfamiliar, so the appropriate course of action is to use one of the pedagogical dictionaries for learners of English, since the other dictionary types (bilingual dictionaries or those designed for native speakers of English) are not reliable sources of this sort of information. After finding the entry for the appropriate word (see Figure 1 above), they should identify the numbered sense that corresponds to the usage in the sentence, select an appropriate pattern from among those provided—in this case, *warn somebody about something* or *of something*—and then return to the task page and make the required changes.

The task, which was used as both a pre-test and post-test, consisted of 10 items presented in randomized order. The sentences were found in the Longman Learners Corpus by searching high-frequency node words featured in pattern-grammar references (Francis, Hunston, & Manning, 1996, 1998). The sentences were edited to make the context clear and to isolate the pattern error as the only infelicity. A list of acceptable corrections was compiled using information from pedagogical dictionaries and the Corpus of Contemporary American English (Davies, 2008), and by surveying eight native speakers of English who have master's degrees or higher in TESOL or applied linguistics and three or more years of ESL/EFL

teaching experience.

Item characteristics are summarized in Table 3. Four items dealt with verb patterns, four with noun patterns, and two with adjective patterns. Most corrections involved changes to prepositions, usually by substitution but in some cases by addition or removal. Some substitutions involved complement clauses: an *-ing* form in the case of one verb pattern, and a *that*-clause in the case of two adjective patterns. In several cases, different types of correction were possible.

**Table 3.** *Pattern Identification and Correction Task Items Characterized by Node Word Type and Type of Correction Required*

Node word type	Number of items	Correction type required	Number of items
Verb	4	substitute complement clause	3*
Noun	4	substitute preposition	7*
Adjective	2	add preposition	1 <sup>†</sup>
		remove preposition	2 <sup>†</sup>

*Note.* 10 items total; \* either correction type was acceptable in the case of two items; <sup>†</sup> either correction type was acceptable in the case of one item.

Polytomous scoring was used to allow differentiation between the ability to identify an error and the ability to correct it. One point was awarded if the participant made any change to the grammatical component of the pattern, that is, to any part other than the lexical node word. Such a change was interpreted as evidence that the error had been located, so the variable yielded by this score was called ID. An additional point was awarded if the change matched an item in the list of acceptable responses, yielding another variable called CORRECT. With 10 points possible for both ID and CORRECT, the maximum score on the task was 20. Pre-test reliability (Cronbach's alpha) for ID was .70 and for CORRECT was .71, while the post-test coefficients were .85 and .83, respectively, indicating consistency between the two measures.

### Dictionary Use Measures

In addition to providing performance scores, the pattern identification and correction task was also the source of process data assessing use of dictionaries. Each participant's onscreen performances on both administrations of the task were video recorded using Camtasia, a screen-capture program. The video files were then imported into Transana, a software tool for analyzing audio and video data, where they were coded and transformed into two quantitative measures of referencing ability:

1. **Lookups.** These were counts of the number of individual dictionary searches performed by each participant. Searches were divided into two categories: *learner-dictionary lookups* (i.e., in monolingual, pedagogical dictionaries) and *non-learner dictionary lookups* (i.e., in bilingual dictionaries, native speaker dictionaries, or non-dictionary tools such as search engines). When screen-capture footage showed a particular word, phrase, or sentence being searched, a lookup was recorded in the appropriate category, except in cases where the same search had already been conducted in the same resource in the same administration of the task.
2. **Lookup points.** These were points achieved on the task that were attributed to use of a learner dictionary, another type of reference, or else to a participant's use of pre-existing knowledge, intuition, or guesswork. If screen-capture footage showed explicit information in an online resource being used in a correct or partially correct response, the resulting points were coded either as *learner dictionary lookup points* or *non-learner dictionary lookup points*, depending on the type of resource. If language data used in a correct or partially correct response could not

be traced to an online resource, those points were attributed to pre-existing knowledge, intuition, or guesswork and coded as *null lookup points*.<sup>3</sup>

## Questionnaires

The online instruments used to elicit perception data were (a) a 3-item post-tutorial questionnaire administered multiple times, immediately after the participants' first completed attempt at each tutorial, and (b) a 25-item end-of-course evaluation. The post-tutorial questionnaire, consisting of closed-ended questions, was designed to capture basic snapshots of perceived levels of interest, challenge, and usefulness of the tutorials throughout the progression of instruction. The end-of-course evaluation, which included both closed- and open-ended items, sought more detailed, summative perceptions of learning achieved through the materials. Both instruments were administered anonymously and separately from those procedures that were screen-captured, to encourage more accurate, less self-protective responses (Dörnyei & Taguchi, 2003).

## RESULTS

### Research Question 1: Effects of Treatment Condition

#### *Pattern Identification and Correction*

Descriptive statistics for ID and CORRECT scores are reported in Table 4, along with combined totals and effect sizes. The averages showed that, across time and conditions, more points were achieved by identifying errors than by correcting them, which makes sense given the nature of the task and scoring method; that is, while every error that is corrected must necessarily be identified, the converse is not true. However, as will be shown, more errors were both identified and corrected than were identified alone.

Table 4. *Descriptive Statistics and Effect Sizes for ID and CORRECT Scores*

	ID		CORRECT		TOTAL (ID + CORRECT)	
	SI	DP	SI	DP	SI	DP
	<i>n</i> = 32	<i>n</i> = 32	<i>n</i> = 32	<i>n</i> = 32	<i>n</i> = 32	<i>n</i> = 32
Pre-test mean	4.13 (2.84)	3.78 (2.01)	2.75 (2.30)	2.25 (1.95)	6.87 (5.05)	6.03 (3.80)
Post-test mean	7.88 (2.50)	4.31 (2.60)	6.00 (2.66)	2.91 (2.53)	13.88 (4.96)	7.22 (4.94)
Mean gain	3.75	.53	3.25	.66	7.01	1.19
Pre-Post <i>d</i>	1.40	.23	1.31	.29	1.40	.27
Between groups <i>d</i> (Post)		1.40		1.19		1.34

*Note.* Maximum score for ID and CORRECT = 10. SI = strategy instruction; DP = dictionary practice. Standard deviations have been placed in parentheses.

At pre-test, the SI group demonstrated a slight edge over the DP group in both ID and CORRECT, as well as more dispersion in those scores, while at post-test the gaps in scores widened substantially and variation decreased. Using either the general index in J. Cohen (1988) or a more stringent scale recently proposed specifically for L2 research (Oswald & Plonsky, 2010),<sup>4</sup> the effect sizes for pre-to-post differences on either the total combined score, or the ID and CORRECT scores individually, would be characterized as large for the SI group and small for the DP group. The post-test, between-group differences for both measures would also be considered large effects.

Correlational analysis showed a strong, positive relationship between ID and CORRECT ( $r = .90$ ,  $p < .001$ ), so a multivariate analysis of variance with these scores as separate dependent variables was not appropriate. Thus the scores were combined and the totals were submitted to a univariate analysis of

variance (ANOVA) using the GLM procedure in SPSS. The ANOVA employed a randomized complete block design, with treatment condition and matched pair (based on VLT scores) as the between-subjects factors, and time as the within-subjects factor. Because time had only two levels, sphericity was not a concern, but analysis of histograms and Q-Q plots indicated non-normality in both the pre- and post-test data. Base-10 logarithmic transformations, which are recommended in cases of substantial positive and negative skew (Tabachnick & Fidell, 2007), failed to resolve these issues, so it was decided to continue with the analysis using the untransformed data, recognizing that this could lead to a loss in statistical power. The ANOVA showed a significant main effect for time,  $F(1,31) = 50.83, p < .001$ , and a significant effect for the interaction of time and group  $F(1,31) = 25.7, p < .001$ , but no effect for the interaction of time and matched pair,  $F(1,31) = 1.11, p < .391$ .

Because the ordinal variable of matched pair represented vocabulary size only indirectly, the influence of this construct was further investigated by regressing the combined total scores (ID + CORRECT) on VLT scores. For the sample as a whole, vocabulary size could explain a small but significant amount of variation on pre-test performance,  $R^2 = .22, p < .001$ . Thus, as the scatterplot in Figure 7 (left side) indicates, a larger vocabulary size did not provide participants with much advantage on the task at pre-test.

At post-test, regression analyses suggested VLT score was a better predictor for the SI group,  $R^2 = .36, p < .001$ , compared to the DP group,  $R^2 = .11, p = .06$ . However, the scatterplot showed two outliers<sup>5</sup> in the SI group falling relatively near the regression line (Figure 7, right side). When these data were removed from the analysis, the results changed considerably,  $R^2 = .07, p = .176$ . Treatment effects, then, were not significantly related to vocabulary size, contrary to findings in the pilot study, which was based on a smaller sample ( $n = 38$ ).

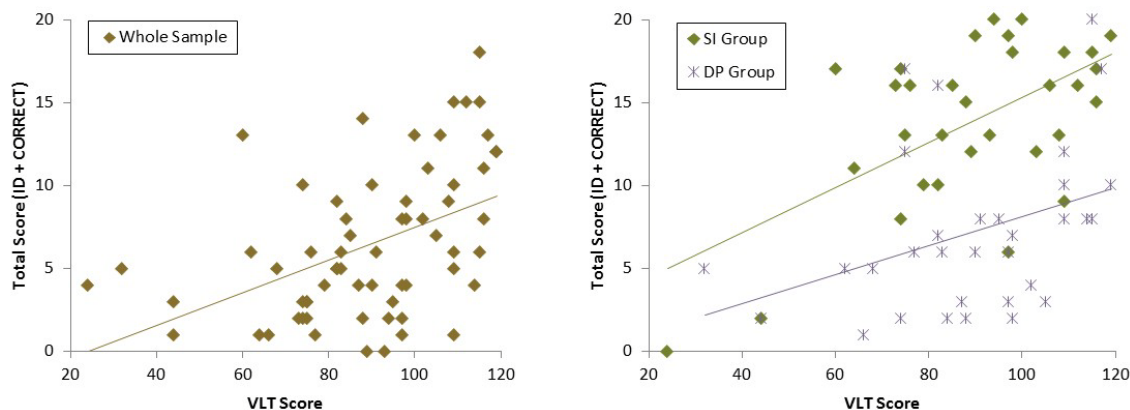


Figure 7. Left side: Relationship between total score (ID + CORRECT) and Vocabulary Levels Test score for whole sample at pre-test ( $N = 64$ ). Right side: Relationships between total score (ID + CORRECT) and Vocabulary Levels Test score at post-test ( $n = 32$  for both groups).

### Dictionary Use Measures

**Lookups.** Descriptive statistics and effects sizes for dictionary lookups are reported in Table 5. At pre-test, the averages were similar, with both groups conducting less than three lookups in learner dictionaries and less than four in non-learner dictionaries. Patterns diverged markedly at post-test, however, with the SI group's learner dictionary lookups increasing to an average of 9.53 (nearly one per item on the task), while among the DP group the increase is less than half a lookup. Use of non-learner dictionaries decreased for both groups, but again the change was more substantial among the SI group, with such lookups nearly dropping off completely. For both categories of lookup, effect sizes for the within-group,



Table 5. Descriptive Statistics and Effect Sizes for Dictionary Lookups

	Learner dictionary lookups		Non-learner dictionary lookups		Total lookups (learner + non-learner)	
	SI	DP	SI	DP	SI	DP
	<i>n</i> = 32	<i>n</i> = 32	<i>n</i> = 32	<i>n</i> = 32	<i>n</i> = 32	<i>n</i> = 32
Pre-test mean	2.63 (3.91)	2.75 (3.94)	3.97 (4.74)	3.69 (4.90)	6.59 (6.22)	6.44 (5.18)
Post-test mean	9.53 (4.60)	3.34 (4.71)	.16 (0.52)	2.84 (3.73)	9.69 (4.70)	6.19 (5.17)
Mean change	6.90	.59	-3.81	-.85	3.10	-.25
Pre-Post <i>d</i>	1.62	.14	-1.13	-.19	.56	-.05
Between groups <i>d</i> (Post)		1.33		1.01		.71

Note. SI = strategy instruction; DP = dictionary practice. Standard deviations have been placed in parentheses.

pre-post differences would again be considered large for the SI group and small for the DP group, while between-group differences at post-test represent large effects.

To test for significance, the data were fitted to a generalized linear model using the logarithm as the link function and the Poisson distribution as the probability distribution, which is appropriate in cases where data represent counts and the number of observations is not fixed (Larson-Hall, 2010). Using the GENLIN procedure in SPSS, a Poisson regression was modeled to represent the mean rate at which participants performed lookups, with learner dictionary lookups as the response variable, group and time as predictors, and VLT scores as a covariate. Non-learner dictionary lookups were not included in the model as this would have violated the assumption of independence and because learner dictionary lookups was the variable of primary interest. A test of the full model against an intercept-only model was statistically significant,  $\chi^2(4) = 49.2, p < .001$ . Pairwise comparisons showed the mean difference in learner dictionary lookups for the SI group on the post-test was significant compared to the DP group on the post-test (mean difference = 6.0, 95% CI 2.44, 9.57,  $p < .001$ ), and compared to the SI group on the pre-test (mean difference = 6.7, 95% CI 3.22, 10.17,  $p < .001$ ).

**Dictionary Lookup Points.** Descriptive statistics and effect sizes for dictionary lookup points are reported in Table 6. Pre-test averages again showed similarities between groups, with the majority of points achieved without the aid of a reference, while learner dictionary lookups accounted for less than two points and non-learner dictionary lookups for about half a point, on average. On the post-test, however, the SI group averages showed a sixfold increase in points from learner dictionary lookups, with

Table 6. Descriptive Statistics and Effect Sizes for Lookup Points

	Learner dictionary lookup points		Non-learner dictionary lookup points		Null lookup points	
	SI	DP	SI	DP	SI	DP
	<i>n</i> = 32	<i>n</i> = 32	<i>n</i> = 32	<i>n</i> = 32	<i>n</i> = 32	<i>n</i> = 32
Pre-test	1.78 (3.63)	1.16 (2.69)	.59 (1.39)	.44 (1.04)	4.50 (3.54)	4.44 (2.72)
Post-test	11.06 (6.02)	2.94 (5.30)	.00 (.00)	.41 (1.01)	2.81 (3.54)	3.88 (2.65)
Mean change	9.28	1.78	-.59	-.03	-1.69	-.56
Pre-Post <i>d</i>	1.87	.42	-.60	-.03	-.48	-.21
Between groups <i>d</i> (Post)		1.43		-.57		-.34

Note. All maximum scores = 20; SI = strategy instruction, DP = dictionary practice; standard deviations are in parentheses.

null lookup points reduced by about one third and non-learner dictionary lookup points dropping off completely. For the DP group, by contrast, learner dictionary lookup points increased and non-learner dictionary and null lookup points decreased by much smaller margins. Large effect sizes for learner dictionary lookup points were observed in the within-group, pre-post difference for the SI group, as well as for the between-group difference at post-test.

To test for significance, the mean change scores (see Table 6) were submitted to a two-sample Hotelling's  $T^2$  test using the R statistical software package. Hotelling's  $T^2$  is a generalization of Student's  $t$  for multivariate data, which is used when there are multiple dependent variables likely to be correlated with each other, and when the independent variable comprises only two groups (Tabachnick & Fidell, 2007). The test showed a significant difference between SI and DP score vectors,  $T^2(3,60) = 12.98, p < .001$ . As a post-hoc measure, the mean change scores for each category of lookup point were submitted to Welch two-sample  $t$ -tests, which returned a statistical result in the case of learner dictionary lookup points,  $t(52.43) = 6.22, 95\% \text{ CI } 5.08, 9.92, p < .001$ .

It is notable that, despite the difference in magnitude, post-test gains for both groups are attributable to increased successful exploitation of learner dictionaries. To determine whether the DP group's gains were statistical, another two-sample Hotelling's  $T^2$  test was run using only this group's learner dictionary lookup points, non-learner dictionary lookup points and null lookup points as dependent variables and time as the independent variable. No difference was found:  $T^2(3,60) = 1.03, p = .385$ .

Regardless, these results merited closer scrutiny. Examination of the raw data showed that 78% of the DP group's post-test learner dictionary lookup points were contributed by only six individuals, whose average total post-test score,  $M = 15.7, SD = 3.1$ , greatly exceeded that of the remainder of the group,  $M = 5.27, SD = 2.6$ . Likewise, these individuals accounted for most of the learner dictionary lookups at post-test,  $M = 10.8, SD = 1.7$ , versus  $M = 1.62, SD = 3.2$  for the rest of the group. These participants appear to have acquired some degree of strategic pattern knowledge independently of the strategy instruction. This finding will be addressed further in the discussion.

Finally, to determine whether the treatments had differential effects on participants' ability to identify errors versus their ability to correct them, the number of responses where errors were simply identified versus those where errors were both identified and corrected were tallied for both groups with reference to dictionary type, and proportions were calculated (Table 7). Overall, items where errors were both identified and corrected were more numerous than those where they were only identified, with the exception of pre-test items answered by the DP group using non-learner dictionaries. These data show

Table 7. *Frequencies and Proportions for Response Type by Category of Reference*

		SI			DP		
		Learner dictionary	Non-learner dictionary	No use of references	Learner dictionary	Non-learner dictionary	No use of references
Pre	Identified only	7	3	34	5	6	38
	Identified and corrected	28	8	55	16	4	52
	Identified only : identified and corrected	.25	.37	.62	.31	1.50	.73
Post	Identified only	33	0	26	8	3	32
	Identified and corrected	161	0	32	43	5	46
	Identified only : identified and corrected	.20	0	.81	.19	.60	.70

Note. SI = strategy instruction; DP = dictionary practice.

again that post-test gains for both groups are attributable to increased use of learner dictionaries and, moreover, that while more items overall were answered correctly by the SI group, the proportions in the learner dictionary category were similar; that is, .205 for the SI group versus .186 for the DP group.

To test whether these proportions were statistically different, the data from the learner dictionary category were fitted to another generalized linear model using GENLIN in SPSS, with the logit as the link function and the binomial as the distribution, which is appropriate for binary data with a fixed number of observations (Larson-Hall, 2010). With response type (i.e., *identified only* versus *identified and corrected*) as the dependent variable and group as the predictor, a pairwise comparison showed the difference to be non-significant,  $p = .818$ . Thus, those DP group participants who discovered the strategy for themselves seemed as adept in both identifying and correcting errors as their SI group counterparts. This finding will also be addressed in the discussion.

### Research Question 2: Participant Perceptions

Cumulative data for the post-tutorial questionnaires are provided in the pie charts in Figure 9, which represent averaged ratings of the perceived levels of interest, challenge, and usefulness of the SI tutorials based on 465 responses. Analysis showed that, on the whole, the tutorials were viewed positively, with a majority of participants rating them either *interesting* or *very interesting* (57%), *challenging* or *very challenging* (54%), and *useful* or *very useful* (66%).

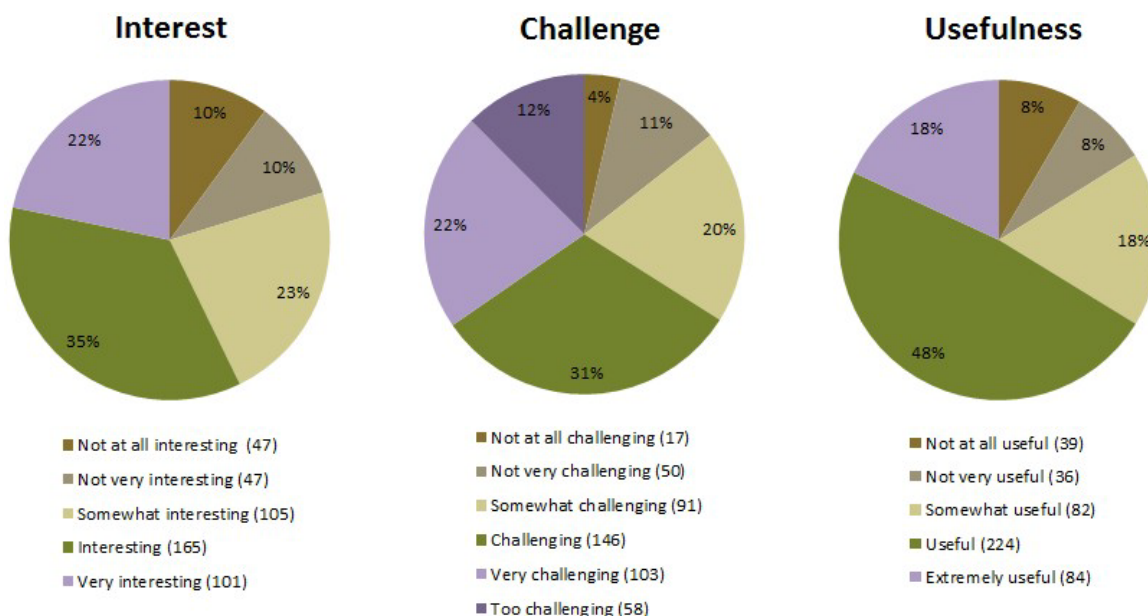


Figure 9. Perception ratings pooled from 465 submitted post-tutorial questionnaires.

Perceptions of learning achieved through the materials were measured by the end-of-course evaluation, in which participants rated statements on a scale from 1 (*strongly disagree*) to 6 (*strongly agree*). Appendix B shows responses to statements regarding the SI materials that have been rank ordered according to the combined percentage of those who responded either *agree* or *strongly agree*. Among the 58 respondents, clear majorities felt the tutorials had improved their understanding of dictionary types (81%) and their ability to exploit a dictionary for word-pattern information (78%), as well as convinced them to use learner dictionaries more often in the future (78%). Most also agreed they had learned things they didn't know about word patterns (86%), had improved their ability to identify word patterns in sentences (84%), and would pay more attention to word patterns in the future while writing in English (84%). However, only a smaller majority said the tutorials had increased their interest in vocabulary learning (59%), while nearly half felt the tutorials had made English vocabulary seem more complicated and difficult (47%).

In addition to Likert-scale items, the end-of-course evaluation included open-ended questions asking what participants liked most about the VVT course, how it could be improved, and why it should or should not be included in the ESL curriculum at Iowa State University. Responses to these questions were imported into Weft QDA, an open-source software tool for qualitative data analysis, where they were coded using an inductive approach to identify common themes. Themes commented on by five or more participants are reported below.

Regarding what they liked most, 10 of the 58 respondents (17%) mentioned the video presentations, nine (15%) mentioned the dictionary training, and five (9%) mentioned learning about word patterns. Example comments included:

I liked the video part. It was easy to understand and learn.

I liked "how to use the dictionary correctly". I know more about how to use the dictionary.  
And the kind of words ... the meaning, usage.

For suggested improvements, 10 participants (17%) mentioned technical issues and design limitations, for example:

Use a user friendly interface, maybe. For example, attractive graphics and design  
... it is better to divide homework into more pieces in order to avoid the operating mistakes on computer which can ruin the homework by the only attempt.<sup>6</sup>

Finally, with respect to curricular integration, three participants (5%) opposed the idea, citing the time required to complete the tutorials and a perceived lack of instructional value, for example:

I do not think VVT is very useful. We spent much time on VVT course. However I felt like I did not learn any useful stuff.

By contrast, 50 participants (86%) were in favor, with perhaps the most articulate statement of support being:

I strongly agree. Actually I even think there should be a class ONLY for vocabulary usage like the VVT Course. This kind of class is essential for English learners. It really helps students to use English properly. Without this kind of knowledge (how to use words correctly and as naturally as native speaker) English learners will end with broken English I believe.

## DISCUSSION

This study investigated the efficacy of a prototype online SI course designed to teach tertiary-level ESL learners the integrated knowledge and skills needed to address a specific vocabulary usage task involving pedagogical dictionaries. Following the intervention, SI group participants displayed significant improvements in their ability to select the correct dictionary for the task, and to exploit that dictionary in both identifying and correcting errors involving lexical patterns, as illustrated in a pair of video excerpts (see [video link 4](#)). In addition, most participants reported that they found the online materials interesting, useful and appropriately challenging, felt their vocabulary knowledge and dictionary skills had benefitted from them, and believed they should be integrated into the ESL curriculum. Together, the findings provide evidence of the potential for some aspects of L2 strategy instruction to be delivered in an automated, online form.

In contrast to SI research generally, the effect sizes documented in this study were large.<sup>7</sup> While the instructional materials and technology-facilitated enhancements to the research design may partly account for these large effects, the sample's generally poor performance on the pre-test must be taken into account, as should the fact that the SI treatment itself was comparatively long, focused on a single set of

behaviors, and in a strategy domain that has been shown to be a more fruitful focus for SI—all of which represent advantages relative to some previous intervention studies (see Plonsky, 2011). Regardless, the findings also demonstrate how technology can be used to increase the validity of SI research.

In addition, the study has addressed calls for integrating dictionary skills training with language awareness, in response to research showing that learners lack conceptual understanding of complementation, transitivity, and grammatical collocation. Indeed, in the absence of strategy instruction, most participants had trouble identifying the majority of errors, and thus made changes to parts of the sentences that were correct, usually creating additional errors in the process, as shown in a video montage (see [video link 5](#)). While some L2 vocabulary experts have proposed addressing such gaps in learners' lexical knowledge through intensive, word-focused instruction (e.g., Laufer, 2005), the present research shows how an approach based on strategies, metalinguistic understanding, and pedagogical references might prove an effective complement, particularly given that strategic gains were seen to accrue largely independently of vocabulary size.

The fact that a small cohort of DP group participants appeared to have developed strategic pattern knowledge through dictionary practice alone, with similar mastery of the subskills involved, calls to mind the “good language learner” in early L2 strategy research (Rubin, 1975). As noted by Tseng and Schmitt (2008), some learners are able to discover effective new strategies on their own through metacognitive monitoring and evaluation of strategy use. Assuming such learners will exist in any population, the question becomes, how to cater to their more advanced needs at the same time as the majority's more fundamental ones?

Automated, online SI represents one possible answer. Because the instruction is self-paced, high ability learners can move quickly through those assignments that represent little challenge for them, while others can take the time they need to gain control over the basic behaviors. This appeared to be the case in the VVT course, as log files showed considerable variation in the amount of time it took learners to reach criterion level on the tutorials. Teachers, meanwhile, can use performance data from the online materials to identify those learners who will need extra help during class, versus those who are ready for more advanced consolidation or extension tasks.

Needless to say, there will be continued need for teachers in the SI process, as there are limits to what can be accomplished through technology alone. Indeed, the greatest limitation of this investigation is that it did not evaluate whether strategic gains transferred to participants' L2 writing or persisted over time. It was assumed, in fact, that in the absence of complementary, teacher-led instruction to facilitate transfer and maintenance, such an effort was unwarranted. Future work should therefore investigate blended approaches that can take best advantage of the respective strengths of computers and teachers, with the work of the latter focused on refinement and contextualized applications of skills developed by the former.

Other limitations of the study suggest focuses for further research more specifically related to pattern grammar and dictionary use. The findings are likely to have been influenced by the predominance of Chinese-speaking learners in the sample. Although this was the population of interest, studies with greater representation of learners from other educational backgrounds, particularly those less skilled in syntactic processing as a result of years of exam-oriented English instruction, are warranted. The contribution of skill acquisition-oriented practice and feedback should also be more directly investigated. For example, eye-tracking could provide finer-grained analyses of treatment effects on dictionary consultation behavior, as screen-capture data showed untrained participants apparently overlooking pattern information even when they managed to find the correct entry (see [video link 6](#)). Finally, case studies could help identify the specific factors that allow individuals, like those in the DP group, to bootstrap their way into strategic pattern knowledge.



## CONCLUSION

As noted, lifting some of the burden of L2 strategy instruction from teachers and timetables could shift the cost-benefit balance toward including it in syllabuses and language programs. While a great deal of attention has been devoted to the search for CALL applications that can provide more effective instructional opportunities for learners, few commensurate efforts have tried to harness the power of technology in enhancing learners' ability to learn. There is clearly scope for increasing such efforts, assuming worthwhile strategic goals that lend themselves to automated, online SI can be identified. The possibilities are limited only by the imaginations of SI researchers, instructional designers and technology-minded teachers.

## APPENDIX A. Screenshot of Online Version of the Vocabulary Levels Test

**2000 Word Level**

1. original 2. private 3. royal 4. slow 5. sorry 6. total	<input type="checkbox"/> complete <input type="checkbox"/> first <input type="checkbox"/> not public
1. blame 2. hide 3. hit 4. invite 5. pour 6. spoil	<input type="checkbox"/> keep away from sight <input type="checkbox"/> have a bad effect on something <input type="checkbox"/> ask
1. basket 2. crop 3. flesh 4. salary 5. temperature 6. thread	<input type="checkbox"/> money paid for regularly doing a job <input type="checkbox"/> heat <input type="checkbox"/> meat

1. apply 2. elect 3. jump 4. manufacture 5. melt 6. threaten	<input type="checkbox"/> choose by voting <input type="checkbox"/> become like water <input type="checkbox"/> make
1. accident 2. choice 3. debt 4. fortune 5. pride 6. roar	<input type="checkbox"/> having a high opinion of yourself <input type="checkbox"/> something you must pay <input type="checkbox"/> loud, deep sound
1. birth 2. dust 3. operation 4. row 5. sport 6. victory	<input type="checkbox"/> being born <input type="checkbox"/> game <input type="checkbox"/> winning

*Adapted from Schmitt et al. (2001)*

## APPENDIX B. Rank Ordering of Participants' Perceptions of Learning Outcomes as Elicited by End-of-Course Evaluation

	Responses of <i>agree or strongly agree</i> ( <i>n</i> = 58)
<b>1. Because of the VVT Tutorials, ...</b>	
... I am able to identify word patterns in sentences more easily.	49 (84%)
... I will pay more attention to word patterns when I write.	49 (84%)
... I have more confidence in my ability to use new vocabulary correctly.	47 (81%)
... I better understand the differences between types of dictionaries (e.g., monolingual, bilingual, native-speaker).	47 (81%)
... I better understand the connection between word usage and word meaning.	45 (78%)
... I am able to find word pattern information in the dictionary more efficiently.	45 (78%)
... I will use online monolingual learner's dictionaries (e.g., Longman) more often.	45 (78%)
... learning English vocabulary now seems more complicated and difficult.	27 (47%)
<b>2. The VVT Tutorials ...</b>	
... taught me things I didn't know about word patterns.	50 (86%)
... should be part of English 101B.	48 (83%)
... helped me use monolingual learner's dictionaries (e.g., Longman) more effectively.	46 (79%)
... taught me the importance of learning more about words I am already familiar with.	44 (76%)
... improved my ability to edit and correct my own writing.	41 (71%)
... were easy to use.	39 (67%)
... made me feel more interested in learning English vocabulary.	34 (59%)

### NOTES

1. While *task* is typically used in second language acquisition to mean a learning activity involving a primary focus on meaning and a lack of restrictions on learners' linguistic choices (see Ellis, 2003), the much broader conceptualization adopted here is "a structured plan for the provision of opportunities for the refinement of knowledge and capabilities entailed in a new language and its use" (Breen, 1989, p. 187).

2. The SI group mean was 89.1 ( $SD = 21.7$ ) and the DP group mean was 89.3 ( $SD = 20.8$ ), out of a possible 120 on the Vocabulary Levels Test. The nonparametric Mann-Whitney U Test, used because of strongly negative skews in both distributions, showed no statistical difference,  $U = 510.5$ ,  $p = .984$ .

3. In 12 cases, it was unclear to the researcher whether information from Chinese-English dictionaries had been used to achieve correct or partially correct responses, so a PhD student in applied linguistics whose first language is Mandarin Chinese was asked to review the relevant video excerpts to make a determination.

4. The benchmarks in J. Cohen (1988) are .20 for a small effect, .50 for a medium effect, and .80 for a large effect, while Oswald and Plonsky (2010) propose .40, .70 and 1.00, respectively.
5. These two, low-performing SI participants had not completed most of the VVT assignments, as well as many assignments in the composition course, and one ended up dropping in the tenth week while the other failed. Although outliers in statistical terms (as was the DP group participant who scored 18 on the pre-test and 20 on the post-test) similar performances had been observed in previous rounds of piloting, and thus these individuals were deemed representative of the target population. Furthermore, because removal of their data did not change the pattern of results in the major statistical tests of group differences, they were retained in the sample.
6. This comment refers to the fact that participants were told certain tutorials had to be completed successfully in a single attempt, although in fact additional attempts were often allowed at the researcher's discretion. This had become necessary after it was discovered that, in tasks involving multiple-choice questions, some participants were selecting responses at random to quickly identify the correct ones, then restarting the activity. Obviously, this is a bug that needs to be remedied.
7. For comparison's sake, the effect size for the change in the treatment group's total combined scores from pre- to post-test, ( $d = 1.4$ ), can be contrasted with the overall meta-analytic effect size for SI studies in Plonsky (2011),  $M(d) = .49$ ,  $K = 95$ . For Plonsky's subsample of SI studies in which vocabulary knowledge was an outcome variable, the effect size was higher,  $M(d) = .64$ ,  $K = 32$ , and for studies in which the outcome variable was a measure of strategy use, as in the present case, it was higher still,  $M(d) = 1.11$ ,  $K = 9$ . However, this latter category included samples in which the measure focused on quantity, i.e. counts, of strategy use. Thus, simple parallels are hard to draw, insofar as the main measure in the present study assessed the quality of strategy use. Another reason why comparisons are difficult is that, as noted by Gu (2003), most SI studies involving vocabulary have addressed memory strategies, such as the keyword technique.

---

## ACKNOWLEDGEMENTS

The author wishes to thank the anonymous reviewers for their insightful comments and suggestions, and to acknowledge the ESL composition teachers at Iowa State University whose cooperation was crucial to the success of the project, particularly Hyejin Yang and Edna Lima.

---

## REFERENCES

- Abraham, R. G., & Vann, R. J. (1996). Using task products to assess second language learning processes. *Applied Language Learning*, 1(2), 61–89.
- Atkins, B. T. S., & Varantola, K. (1998). Language learners using dictionaries: The final report of the EURALEX/AILA research project on dictionary use. In B. T. S. Atkins (Ed.), *Using dictionaries: Studies of dictionary use by language learners and translators* (pp. 21–82). Tübingen, Germany: Max Niemeyer Verlag GmbH.
- Bogaards, P., & van der Kloot, W. A. (2002). Verb constructions in learners' dictionaries. In A. Braasch & C. Povslen (Eds.), *Proceedings of the Tenth Euralex International Congress, Euralex 2002, Copenhagen, Vol. 2* (pp. 747–757). Copenhagen, Denmark: Center for Sprogteknologi.
- Breen, M. (1989). The evaluation cycle for language learning tasks. In R. K. Johnson (Ed.), *The second language curriculum* (pp. 187–206). Cambridge, UK: Cambridge University Press.

- Carlson, R. (2003). Skill learning. In L. Nadel (Ed.), *Encyclopedia of cognitive science* (pp. 36–42). London, UK: Macmillan.
- Chamot, A. U. (2005). Language learning strategy instruction: Current issues and research. *Annual Review of Applied Linguistics*, 25, 112–130.
- Chan, A. Y. W. (2010). Toward a Taxonomy of Written Errors: Investigation Into the Written Errors of Hong Kong Cantonese ESL Learners. *TESOL Quarterly*, 44(2), 295–319.
- Chan, A. Y. W. (2011). The use of a monolingual dictionary for meaning determination by advanced Cantonese ESL learners in Hong Kong. *Applied Linguistics*. doi: 10.1093/applin/amr038
- Chan, A. Y. W. (2012). Cantonese ESL learners' use of grammatical information in a monolingual dictionary for determining the correct use of a target word. *International Journal of Lexicography*, 25(1), 68–94. doi: 10.1093/ijl/ecr014
- Chapelle, C. A., & Mizuno, S. (1989). Students' strategies with learner-controlled CALL. *CALICO Journal*, 7(2), 25–47.
- Cohen, A. D. (1998). *Strategies in learning and using a second language*. New York, NY: Longman.
- Cohen, A. D. (2007). Becoming a strategic language learner in CALL. *Applied Language Learning*, 17(1–2), 57–71.
- Cohen, A. D. & Ishihara, N. (2005). A web-based approach to strategic learning of speech acts. Minneapolis, MN: Center for Advanced Research on Language Acquisition (CARLA), University of Minnesota. Retrieved from <http://www.carla.umn.edu/speechacts/Japanese%20Speech%20Act%20Report%20Rev.%20June05.pdf>
- Cohen, A. D., & Macaro, E. (2007). *Language learner strategies: Thirty years of research and practice*. Oxford, UK: Oxford University Press.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences (2nd edition)*. Hillsdale, NJ: Erlbaum.
- Coxhead, A. (1998). *An academic word list (English Language Institute Occasional Publication No. 18)*. Wellington, New Zealand: Victoria University of Wellington.
- Davies, M. (2008). The Corpus of Contemporary American English: 425 million words, 1990–present. Available online at <http://corpus.byu.edu/coca/>
- DeKeyser, R. M. (2007). Skill acquisition theory. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition* (pp. 97–114). Mahwah, NJ: Lawrence Erlbaum Associates.
- Dörnyei, Z., & Taguchi, T. (2003). *Questionnaires in second language research: Construction, administration, and processing*. New York, NY: Routledge.
- Dreyer, C., & Nel, C. (2003). Teaching reading strategies and reading comprehension within a technology-enhanced learning environment. *System*, 31(3), 349–365. doi: 10.1016/s0346-251x(03)00047-2
- Dziemianko, A. (2006). *User-friendliness of verb syntax in pedagogical dictionaries of English*. Tübingen, Germany: Max Niemeyer Verlag.
- Dziemianko, A. (2010). Paper or electronic? The role of dictionary form in language reception, production and the retention of meaning and collocations. *International Journal of Lexicography*, 23(3), 257–273. doi: 10.1093/ijl/ecp040

- Dziemianko, A. (2011). User-friendliness of noun and verb coding systems in pedagogical dictionaries of English: A case of Polish learners. *International Journal of Lexicography*, 24(1), 50–78. doi: 10.1093/ijl/ecq037
- Ellis, R. (2003). *Task-based language learning and teaching*. Oxford, UK: Oxford University Press.
- English Department. (n.d.). Placement exam scoring guide. Unpublished scoring guide, Iowa State University.
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G\*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39, 175–191.
- Francis, G., Hunston, S., & Manning, E. (1996). *Grammar Patterns 1: Verbs*. New York, NY: HarperCollins.
- Francis, G., Hunston, S., & Manning, E. (Eds.). (1998). *Grammar patterns 2: Nouns and Adjectives*. London, UK: Harper Collins.
- Frankenberg-Garcia, A. (2005). A peek into what today's language learners as researchers actually do. *International Journal of Lexicography*, 18(3), 335–355.
- Frankenberg-Garcia, A. (2011). Beyond L1-L2 equivalents: Where do users of English as a foreign language turn for help? *International Journal of Lexicography*, 24(1), 97–123.
- Granger, S. (1998). Prefabricated patterns in advanced EFL writing: Collocations and formulae. In A. P. Cowie (Ed.), *Phraseology: Theory, analysis, and applications* (pp. 145–160). Oxford, UK: Oxford University Press.
- Gu, P. Y. (2003). Vocabulary learning in a second language: Person, task, context and strategies. *TESL-EJ*, 7(2), 1–25.
- Harris, V. (2003). Adapting classroom-based strategy instruction to a distance learning context. *TESL-EJ*, 7(2), 1–19.
- Hassan, X., Macaro, E., Mason, D., Nye, G., Smith, P., & Vanderplank, R. (2005). Strategy training in language learning: A systematic review of available research. *Research Evidence in Education Library*. Downloaded from <http://eppi.ioe.ac.uk/cms/Default.aspx?tabid=296>
- Hauck, M., & Hampel, R. (2008). Strategies for online learning environments. In S. Hurd & T. Lewis (Eds.), *Language learning strategies in independent settings* (pp. 283–302). Bristol, UK: Multilingual Matters.
- Hemchua, S., & Schmitt, N. (2006). An analysis of lexical errors in the English compositions of Thai learners. *Prospect*, 21(3), 3–25.
- Huang, H.-C., Chern, C.-L., & Lin, C.-C. (2009). EFL learners' use of online reading strategies and comprehension of texts: An exploratory study. *Computers & Education*, 52(1), 13–26. doi: 10.1016/j.compedu.2008.06.003
- Hubbard, P. (2004). Learner training for effective use of CALL. In S. Fotos & C. Browne (Eds.), *New perspectives on CALL for second language classrooms* (pp. 45–68). Mahwah, NJ: Lawrence Erlbaum Associates.
- Hunston, S., & Francis, G. (1998). Verbs observed: A corpus-driven pedagogic grammar. *Applied Linguistics*, 19(1), 45–72.
- Hunston, S., & Francis, G. (2000). *Pattern grammar: A corpus-driven approach to the lexical grammar of English*. Amsterdam, Netherlands: John Benjamins.



- Hunston, S., Francis, G., & Manning, E. (1997). Grammar and vocabulary: Showing the connections. *ELT Journal*, 51(3), 208–216.
- Jamieson, J., & Chapelle, C. A. (1987). Working styles on computers as evidence of second language learning strategies. *Language Learning*, 37(4), 523–543.
- Kim, K. J., Bonk, C. J., & Zeng, T. T. (2005). Surveying the future of workplace e-learning: The rise of blending, interactivity, and authentic learning. *eLearn*, 2005(6), 2.
- Kohler, D. B. (2002). *The effects of metacognitive language learning strategy training on lower-achieving second language learners*. Unpublished PhD dissertation, Brigham Young University.
- Larson-Hall, J. (2010). *A guide to doing statistics in second language research using SPSS*. New York, NY: Routledge.
- Laufer, B. (2005). Focus on form in second language vocabulary learning. In S. H. Foster-Cohen, M. del Pilar García Mayo & J. Cenoz (Eds.), *EUROSLA Yearbook* (Vol. 5, pp. 223–250). Amsterdam, Netherlands: John Benjamins.
- Laufer, B. (2011). The contribution of dictionary use to the production and retention of collocations in a second language. *International Journal of Lexicography*, 24(1), 29–49.
- Lew, R. (2011). Studies in dictionary use: Recent developments. *International Journal of Lexicography*, 24(1), 1–4.
- Li, J. (2009). The evolution of vocabulary learning strategies in a computer-mediated reading environment. *CALICO Journal*, 27(1), 118–146.
- Nation, I. S. P. (1990). *Teaching and learning vocabulary*. Boston, MA: Newbury House.
- Nesi, H. (2000). *The use and abuse of EFL dictionaries*. Tübingen, Germany: Max Niemeyer Verlag GmbH.
- Nesi, H., & Haill, R. (2002). A study of dictionary use by international students at a British university. *International Journal of Lexicography*, 15(4), 277–305. doi: 10.1093/ijl/15.4.277
- Nesi, H., & Tan, K. H. (2011). The effect of menus and signposting on the speed and accuracy of sense selection. *International Journal of Lexicography*, 24(1), 79–96. doi: 10.1093/ijl/ecq040
- Neubach, A., & Cohen, A. D. (1988). Processing strategies and problems encountered in the use of dictionaries. *Dictionaries: Journal of the Dictionary Society of North America*, 10, 1–19.
- O'Bryan, A., & Hegelheimer, V. (2007). Integrating CALL into the classroom: The role of podcasting in an ESL listening strategies course. *ReCALL*, 19(2), 162–180. doi: 10.1017/s0958344007000523
- O'Malley, J. M., & Chamot, A. U. (1990). *Learning strategies in second language acquisition*. Cambridge, UK: Cambridge University Press.
- Oswald, F. L., & Plonsky, L. (2010). Meta-analysis in second language research: Choices and challenges. *Annual Review of Applied Linguistics*, 30, 85–110. doi: 10.1017/S0267190510000115
- Oxford, R. L., Cho, Y., Leung, S., & Kim, H.-J. (2004). Effect of the presence and difficulty of task on strategy use: An exploratory study. *International Review of Applied Linguistics*, 42(1), 1–47.
- Plonsky, L. (2011). The effectiveness of second language strategy instruction: A meta-analysis. *Language Learning*, 61(4), 993–1038. doi: 10.1111/j.1467-9922.2011.00663.x
- Pujolá, J. T. (2002). CALLing for help: Researching language learning strategies using help facilities in a web-based multimedia program. *ReCALL*, 14(02), 235–262.

- Ranalli, J. (2013). Designing online strategy instruction for integrated vocabulary depth of knowledge and web-based dictionary skills. *CALICO Journal*, 30(1), 16–43.
- Read, J. (2004). Plumbing the depths: How should the construct of vocabulary knowledge be defined? In P. Bogaards & B. Laufer (Eds.), *Vocabulary in a second language: Selection, acquisition and testing* (pp. 209–227). Amsterdam, Netherlands: John Benjamins.
- Rees–Miller, J. (1993). A critical appraisal of learner training: Theoretical bases and teaching implications. *TESOL Quarterly*, 27(4), 679–689.
- Reinders, H., & White, C. (2010). The theory and practice of technology in materials development and task design. In N. Harwood (Ed.), *English language teaching materials: Theory and practice* (pp. 58–80). Cambridge, UK: Cambridge University Press.
- Rubin, J. (1975). What the "good language learner" can teach us. *TESOL Quarterly*, 9, 41–51.
- Rubin, J., Chamot, A. U., Harris, V., & Anderson, N. J. (2007). Intervening in the use of strategies. In A. D. Cohen & E. Macaro (Eds.), *Language learner strategies: Thirty years of research and practice* (pp. 141–160). Oxford, UK: Oxford University Press.
- Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the vocabulary levels test. *Language testing*, 18(1), 55–88.
- Sykes, J. M., & Cohen, A. D. (2008). *Observed learner behavior, reported use, and evaluation of a website for learning Spanish pragmatics*. Paper presented at the Selected Proceedings of the 2007 Second Language Research Forum, Somerville, MA.
- Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics (5th edition)*. Boston, MA: Allyn and Bacon.
- Tono, Y. (2001). *Research on dictionary use in the context of foreign language learning*. Tübingen, Germany: Max Niemeyer Verlag GmbH.
- Tono, Y. (2011). Application of eye-tracking in EFL learners' dictionary look-up process research. *International Journal of Lexicography*, 24(1), 124–153.
- Tseng, W. T., & Schmitt, N. (2008). Toward a model of motivated vocabulary learning: A structural equation modeling approach. *Language Learning*, 58(2), 357–400.
- Ulitsky, H. (2000). Language learner strategies with technology. *Journal of Educational Computing Research*, 22(3), 285–322.
- Wenden, A. L. (1991). *Learner strategies for learner autonomy*. Hemel Hempstead, UK: Prentice Hall.
- Wenden, A. L. (1995). Learner training in context: A knowledge-based approach. *System*, 23(2), 183–194. doi: 10.1016/0346-251X(95)00007-7
- Winke, P. M., & Goertler, S. (2008). Did we forget someone? Students' computer access and literacy for CALL. *CALICO Journal*, 25(3), 482–509.
- Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge, UK: Cambridge University Press.